EH

faculteit der economische wetenschappen

RESEARCH MEMORANDUM

TILBURG UNIVERSITY

DEPARTMENT OF ECONOMICS

Postbus 90153 - 5000 LE Tilburg
Netherlands

FEW
127

STATISTICAL ANALYSIS OF STEADY-STATE SIMULATIONS:

SURVEY OF RECENT PROGRESS

Jack P.C. Kleijnen
Tilburg University
Tilburg, Netherlands

127

February 1983

ABSTRACT

For non-terminating simulations with steady-state responses the follow-
ing alternative analysis techniques are evaluated: (i) Replicated runs.
(ii) Approximately independent subruns or batches. (iii) Renewal or
regenerative analysis. Two general techniques for reducing nonnormality
and bias are grouping (or batching) and jackknifing respectively. Appli-
cations of the various techniques are discussed. Besides the estimation
of the mean, the estimation of variances and quantiles is presented.

1. Introduction

In this paper we survey the state-of-the-art in statistical analysis of stochastic output of simulation models with steady-state responses. In other words we exclude terminating systems such as queuing systems that close down at the end of the day; see Kleijnen (1975, 1984) for a discussion of terminating versus non-terminating, steady-state systems. We also exclude the analysis of simulation runs corresponding to many different inputs (for which we recommend regression analysis and experimental designs).

Our survey is meant for simulation practitioners with a basic knowledge of statistics. We concentrate on statistical techniques that are of major practical use and that have been developed in the last decade. In practice steady-state ouput analysis is necessary to reveal the limitations of conclusions based on running a simulation model, i.e., if the run is too short the output's accuracy - as measured by a statistical confidence interval - may be very bad. We exclude the statistical problems arising when the accuracy is found to be too low so that the analyst must determine how much longer to run his simulation; runlength determination is discussed at length in Kleijnen (1975, 1984).

So in the remainder we concentrate on the following problem: we perform a simulation experiment with a single variant of a random simulation model, i.e., in the simulation program we fix all input data and all mathematical relationships and we run this program with a sequence of random numbers resulting in a time series. From this time series we

compute "the" response or output, e.g., the average output or the 90% quantile (value exceeded with 10% chance). If "the" response were computed from n independent observations — instead of a time series — then classical statistics would yield the necessary confidence intervals, for instance, for the true population mean μ we could use the t statistic:

$$P(\bar{x} - t_{n-1}^{\alpha/2} s_x / \sqrt{n} < \mu < \bar{x} + t_{n-1}^{\alpha/2} s_x / \sqrt{n}) = 1 - \alpha \qquad (1)$$

In dynamic simulation, however, the output forms a time series. In this paper we survey relevant techniques, and we examine their assumptions. We also report on practical experience with these techniques in a simulation context.

2. Elementary steady-state concepts

In nonterminating simulations there is no critical event that stops the simulation run. For example, we may use simulation to check whether an analytical approximation to a specific queuing problem is not too crude. Typically, in nonterminating simulations we seek the response in the steady-state. Let us examine the steady-state concept in more detail.

Consider a sequence of variables $x_1, x_2, \ldots, x_T$ (or a time series x_t with $t = 1, \ldots, T$). This sequence is strictly stationary (or in its steady state) if the joint distribution function $F_x(x_1, \ldots, x_t, \ldots, x_T)$ does not depend on the time index t. A constant joint distribution implies that the marginal distributions $F_t(x_t)$ are the same for each t.
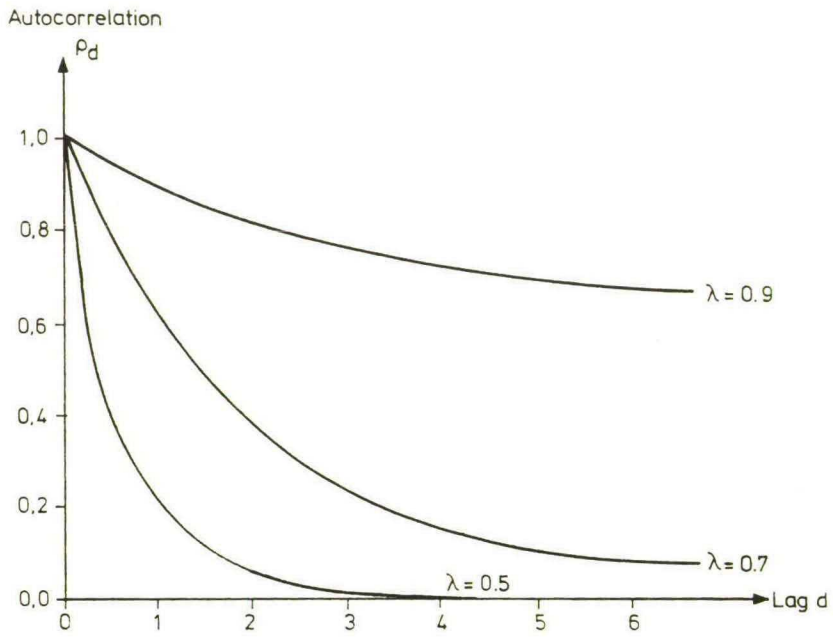
FIG. 1. Autocorrelation ρ_d versus lag d and traffic intensity λ, in an M/M/I queuing system.

Consequently all moments are constant over time, in particular the means and variances: $\mu_t = \mu$ and $\sigma_t^2 = \sigma^2$. Also, the covariance between x_t and x_{t+d} and the correlation ρ_d do not vary wit the absolute point of time t but only with the distance or lag d. The time series is called r-dependent if x_t and $x_{t'}$ are dependent only if their distance $d = |t - t'|$ is not larger than a constant r.

Next we consider the <u>average</u> of a stationary time series: $\bar{x} = \sum_{1}^{T} x_t/T$. Obviously its expectation is μ. It is easy to derive the variance of the average:

$$\text{var}(\bar{x}) = \frac{\sigma^2}{T} \left[1 + 2 \sum_{d=1}^{T} (1 - \frac{d}{T}) \cdot \rho_d \right]. \tag{2}$$

Note that if the observations were independent then $\rho_d = 0$ for $d \geq 1$ and (2) would reduce to the familiar expression $\text{var}(\bar{x}) = \sigma^2/T$. Let us next consider the effects of autocorrelation in a type of systems often simulated, namely queuing systems.

In <u>queuing</u> systems the autocorrelations are positive, i.e., if customer i has to wait relatively long then the next customer probably has to wait long too: $P(w_{i+1} > \mu \mid w_i > \mu) > P(w_{i+1} > \mu)$. These positive correlations inflate the variance of the average; see (2). More specifically - also see Fig. 1 - for M/M/1 queuing systems it can be shown that the (positive) correlations decrease exponentially with the lag d; also these autocorrelations increase with the traffic intensity λ (usually the traffic intensity is denoted by ρ in queuing texts but the symbol ρ is used to denote autocorrelation in statistics texts). The

autocorrelation structure in single-server systems with general arrival and service times is characterized in Woodside and Pagurek (1979). The autocorrelations of M/M/1 systems result in an "inflation factor", i.e., the expression in the square brackets of (2) is as large as 360 when the traffic intensity λ is 0.90; that factor is still 10 when λ is 0.50. So the estimated standard error of the average is completely wrong when the autocorrelations are ignored, as is often done by practitioners and by standard output procedures in many simulation languages! If the autocorrelations are taken into account, then it may turn out that extremely many customers must be simulated. For instance, in an M/M/1 system with traffic load $\lambda = 0.9$ we need to simulate 111,716 customers when we wish to estimate the mean waiting time within one unit with confidence level 0.95; see Fishman (1978b, p. 521).

Since many statistical techniques assume normality, we mention the existence of the stationary r-dependent central limit theorem: given an r-dependent strictly stationary sample x_t (t = 1,...,T) with mean $E(x_t) = \mu$ — and the technical condition that the third absolute moment $E(|x_t|^3)$ exists — the sample mean $\bar{x} = \Sigma x_t/T$ is asymptotically normally distributed; of course the mean of $\bar{x}$ is μ and its variance is given by (2) substituting $\rho_d = 0$ for d > r. Selected references to stationary time series are given in Fishman (1978b), Janssens (1982), Law and Kelton (1982), Mihram (1972).

3. Replicated runs

We can try to analyze nonterminating simulations using the same techniques as used for terminating simulations, the latter type yielding one independent observation per run; see Kleijnen (1975). So suppose we wish to estimate the mean waiting time in the steady state, that is, $\mu = E(w_i)$ for $i \to \infty$. Then we can start with a very long simulation run comprising N customers, and we obtain a single observation on the steady-state response. Denote that first observation by $\bar{w}_1$:

$$\bar{w}_1 = \sum_1^N w_i/N \tag{3}$$

To obtain the next observation $(\bar{w}_2)$ we start all over again, using a different random number stream, so that the next observation on $\bar{w}$ is statistically independent of the previous observation $(\bar{w}_1)$. And so on. This approach would permit a statistical analysis exactly analogous to the analysis for terminating systems, were it not for the following problems.

Since we are interested in the steady-state response the tran-sient (initial, start-up) response creates complications (in a terminating system, a possible transient behavior forms part of "the" response per run; see Kleijnen, 1975, 1984). Hence two options are available:

a. Retain the transient phase.

Though the initial phase creates bias - $E(w_1) < \mu$ - this phase does contain information. Hence it is very well possible that the Mean Squared Error (MSE) is minimized, if the whole time series is used.

Indeed for simple queuing systems it can be proved that the MSE is minimized when the whole run is utilized (assuming the system started in the empty state and the run is long); see Law (1982). Moreover it may be convenient to eliminate worries about the exact length of the transient phase (see below) by simply retaining the whole time series. However, even if the MSE would be minimized, the resulting confidence interval may be inconsistent (wrong confidence level $1 - \alpha$). Actually if we make many replications (n) of a relatively short run (small N), then we obtain a narrow confidence interval around the wrong quantity, i.e. the actual type I error exceeds the nominal α value which is called a "low coverage" of the confidence interval; also see Fishman (1978b) and Law (1982).

Recently Adlakha and Fishman (1982) proposed to start data collection not in the idle state of a queuing simulation but in a con-gested state. If indeed the latter initial state creates bias, we expect that the mean waiting time estimate $\bar{w}$ exceeds the steady-state mean μ. It can be made plausible that the averaged estimate ($\bar{\bar{w}} = \sum_{1}^{n} \bar{w}_j/n$) and the estimated variance ($\widehat{var}(\bar{w})$) are positively correlated; see Kleijnen (1984). Hence the congested initial state is expected to result in an overestimate of $var(\bar{w})$. The latter overestimate may correct the low coverage found in many simulation experiments; see Section 8. Note that though data collection is not started in the empty state, it may be convenient to start the simulation run in the idle state but to defer output generation until a congested state is reached. Wilson and Prits-ker (1978b) experimented with several starting conditions and found that coverage is best when the mode of the steady-state distribution is

selected as initial condition. Of course in practice that mode is un-known but the analyst may try an educated guess. Anyhow these results suggest that the empty state is not the best starting point for data collection if runs are replicated.

b. Eliminate the transient phase.

Practitioners often throw away the initial part of the time series, i.e., the simulation "warms up" before the simulation program starts making observations. Unfortunately two practical problems remain:

(i) How can we determine whether the transient phase is over?

(ii) Throwing away the initial phase of each run wastes computer time.

Sub (i): Practitioners often construct graphs - and making graphs is always an excellent idea in any statistical experiment - to see whether start-up effects "obviously" have disappeared. A guideline may be: "the transient phase is not over as long as the individual waiting times w_i ($i = 1,2,...$) keep growing". Note that running averages like $\sum_i^N w_t/N$ with $N = 1,2,...$ lag behind the individual waiting times. Another simple heuristic is: "throw away the first hundred waiting times". But with heavily congested systems a hundred is too small; remember Fig. 1. Heuristics up to 1978 can be found in Kleijnen (1975) and Wilson and Pritsker (1978a). Recently several statistical techniques for detecting initialization effects have been proposed; see Kleijnen (1984) and Law (1982). The most attractive technique is due to Schruben et al. (1980) and is summarized in Appendix 1. These authors applied their technique to five different simulation models and found that their test is valid and powerful.

Sub (ii): If we overestimate the length of the transient phase we throw away information on the steady-state, and this increases the variance of the final estimator. If we do not wish to waste computer time, we may be tempted to underestimate the initial phase, and we bias the final estimator. Fortunately, if the total time path is long we may assume that the bias caused by the transient phase is negligible. Therefore other approaches besides simple replication have been devised.

4. "Independent" subruns of fixed length: batching

Instead of replicating each long run (with different random numbers) we may make a single, extremely long run. The initial part of that single run may be thrown away; see the preceding section. The remaining (much larger) part of the run is divided into a number of subruns or batches. If the total run is very long, then the subruns will be long, e.g., if the total run (after elimination of the transient phase) comprises 100,000 customers then dividing that run into ten subruns means that each subrun contains 10,000 customers. Now consider subruns 1 and 2: customer 10,000 will affect customer 10,001 but customer 10,000 probably has little effect on customer 11,000; see Fig. . More generally, the last "few" customers of subrun i (i = 1,2,...) do affect the first "few" customers in the next subrun (i + 1) but they hardly affect most customers in the latter subrun. Consequently "the" responses (the averages when estimating μ) of subruns i and i + 1 are practically speaking independent; also see Kleijnen (1975, p. 458). A statistical refinement implies that the independence of the subrun responses is tested (see below); when the hypothesis of independence is

rejected, the subrun length is increased; etc. The subrun length resulting from this statistical refinement, yields n subruns, and each subrun results in a subrun response x_i (i = 1,...,n) where x_i is the subrun average if we want to estimate the steady-state mean μ. (However, in a next section we shall see that we may wish to estimate the pth quantile of the steady-state response and then x_i denotes the pth quantile estimated from subrun i). The techniques for independent observations can now be applied to the (approximately) independent subrun responses x_i (i = 1,...,n): see (1). Next we shall examine some details of the batching approach.

Practitioners often apply a simplified version of the subrun procedure, i.e., they pick an intuitively fixed subrun length and they do not test whether this length indeed yields independent subrun responses x. Such an approach is dangerous because analytical results for simple queuing systems have demonstrated that individual waiting times remain autocorrelated over surprisingly long intervals. Consequently, too short subruns may result in a drastic underestimation of the true variance $var(\bar{x})$; see the comment on Fig. 1. On the other side, if the subruns are unnecessarily long, then only a few subrun responses x remain and the resulting confidence interval tends to become longer and less stable. (Briefly, the interval's expected length increases when the number of batches decreases, primarily because in eq. 1 t_{n-1}^{α} and $1/\sqrt{n}$ increase; the interval's standard error increases primarily because of the lower n value: $var(\hat{\sigma}^2) = 2\sigma^4/(n-1)$. Secondary effects are derived by Schmeiser, 1982.) If we would take the number of batches at its minimum (n = 2) then the coverage could still be $1-\alpha$; however the

interval's expected length and variability would be unacceptable. Schmeiser (1982) recommends the use of ten to twenty batches (or replications if the approach of Section 3 is followed) when computing the confidence interval for the mean μ : $10 \le n \le 20$. However, before we compute that confidence interval we have to "know" that the subrun responses x are independent. To test this independence we certainly need more than twenty subruns, as we shall see next.

The independence of random variables can be tested in many ways. Simulation practitioners may be familiar with the independence issue, in the context of pseudorandom-number generation, i.e., to test whether the pseudorandom numbers r_t (t = 1,2,...,T) are independent many tests have been devised; see recent textbooks like Fishman (1978b) and Law and Kelton (1982). However, in that context extremely many observations are available (T $\rightarrow$ ∞). With subruns the number of observations is much smaller: x_i with i = 1,...,n. The practitioner may be tempted to estimate the autocorrelation between adjacent subrun averages: ρ_1. (Note that in eq. 2 ρ denoted the autocorrelation among individual responses whereas now ρ refers to subrun responses.) However, the usual estimator $\hat{\rho}_1$ is biased and shows a high standard error. We recommend the Von Neumann statistic, say q:

$$q = \sum_{i=1}^{n-1} (x_i - x_{i+1})^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{4}$$

The statistic q concentrates on the first-order autocorrelation ρ_1: if the variables x are independent (implying $\rho_1 = \rho_2 = \cdots = \rho_n = 0$) then, whatever the distribution of x, we have E(q) = 2. However if the vari-

ables x show positive first-order autocorrelation, then $E(q) < 2$. When the variables x are normally and independently distributed (NID) then $\sigma_q^2 = 4(n-2)/\{(n-1)\cdot(n+1)\}$ and the distribution of q is approximately normal for $n > 20$. Consequently we reject the hypothesis of independence if q is smaller than $2-z^{\alpha}\sigma_q$ There is a chance β that we erroneously fail to reject the null hypothesis. This β increases as the autocorrelation ρ_1 is closer to zero and as the sample size n is smaller so that σ_q is larger; based on analytical and Monte Carlo results we recommend to test the independence of the subrun responses x using at least a hundred subruns: $n \geq 100$; see Kleijnen et al. (1982).

Although we recommend the use of at least a hundred subruns when testing the independence of the subrun responses, the final confidence interval for the expected response may be computed from fewer- and hence longer- subruns, so that we obtain extra protection against any dependence not detected in the original (say, hundred) subrun responses. So it may be good practice to compute confidence intervals from only ten to twenty subruns, as proposed by Schmeiser (1982).

If we detect dependence, then we increase the subrun length. The corresponding number of subruns decreases. If the number of subruns becomes smaller than 100 then we need to continue the original simulation run; else we immediately return to the Von Neumann test, etc. The literature gives a number of applications of the subrun approach. Most practical studies do not test the independence of the subrun responses. The oldest procedure including steps _similar_ to our proposal, was derived by Mechanic and McKay in 1966 and was adapted by Fishman (1978 a,

b). We shall return to applications later on.

5. Independent cycles: renewal or regenerative approach

Whereas the subrun approach cuts the total run into pieces such that each piece has the _same_ length, the renewal or regenerative approach cuts the total run into pieces such that the length of a piece is a random variable. Whereas the subrun approach results in responses which are approximately independent, the renewal approach results in perfectly independent responses. First we give an example illustrating the renewal approach.

Consider a simple queuing example, namely the M/M/1 system. Assume we started the simulation with an empty system. Now and then a simulated customer will find the system empty again upon his arrival. The "next history", i.e., the time path of w_t (t = 1,2,...,N) once _a_ customer arrives into an empty system, is completely independent of the past history. This may be illustrated as follows. As soon as a customer, say customer 10, arrives and finds the system empty, we may _erase_ that part of the computer memory containing historical data. Of course we cannot erase the simulation program itself, i.e. for the M/M/1 system we cannot erase the computer instructions executing the following equation where s_i denotes the service time of customer i and a_i denotes the interarrival time between customers i and i − 1:

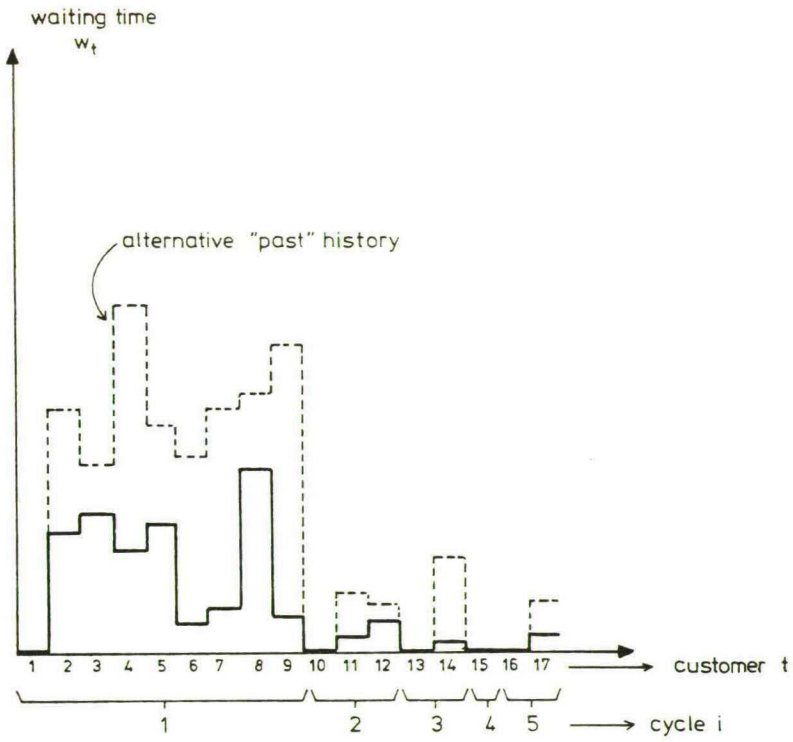$$w_i = \max(w_{i-1} + s_{i-1} + a_i, 0) \qquad i = 2,3,\ldots \qquad (5)$$

FIG. 2. Renewal property.

For the M/M/1 system the historical data consist of the preceding service times s_i and interarrival times a_i (plus the resulting waiting times) for $i < 10$. After we have erased these historical data, we can still compute w_{10}, w_{11}, etc.: by assumption $w_{10} = 0$ (customer 10 arrives into an empty system); using (5) we compute $w_{11} = \max(w_{10}+s_{10}-a_{10},0)$ where s_{10} is sampled from an exponential distribution and a_{10} is sampled from another exponential distribution. Once we know w_{11} we can compute w_{12}, etc. We emphasize that we do not need any historical information when we sample the service and interarrival times, because Poisson processes (resulting in exponential distributions) are memoryless. Fig. 2 further illustrates that the "past history" does not matter. This figure also illustrates the resulting "subruns", called cycles, epochs, blocks, or tours. Each cycle starts as soon as a customer arrives into an empty system. Note when two consecutive customers find the system empty (customers 15 and 16), then two cycles result (cycles 4 and 5). Each cycle begins with a busy period (a customer arrives into an empty system whereupon the system immediately starts serving this customer) and the cycle ends with an idle period (the next cycle starts because a new customer finds the system empty, i.e., the server was idle). In contrast to the batching approach the present analysis results in perfectly independent cycle responses x. The length of the cycles, say L, is a random variable, e.g., in Fig. 2 $P(L_5 > 2) = P(w_{18} > 0)$.

In the above M/M/1 example the empty state was the "renewal" state, i.e., the state starting a new cycle. Crane and Lemoine (1977) give several more examples of renewal states for queuing and inventory systems. In general, all <u>Markov</u> systems have the renewal property. In

such a system the probability of moving to a specific state during period t depends only on the state the system was in at the beginning of that period (and does not depend on previous periods). For instance, in an M/M/s queuing system the system state is specified by the number of customers waiting at each of the s servers. _Any_ state may then be selected as the renewal state! A practical issue is that we wish to select a renewal state such that many cycles result (also see the statistical analysis later on). The renewal state may occur so infrequently that only a few cycles result. For instance, if the queuing system has heavy traffic then the empty state occurs rarely. And some systems have so many possible states that the realization of one particular state occurs rarely. A practical solution is to define a _set_ of states as the _approximate_ renewal state, e.g., the system is "nearly empty", say, in all servers either zero or one customer is waiting. We may test whether this approximate renewal state indeed results in cycle responses which are practically speaking independent. Approximate renewal analysis is discussed in Fishman (1978b), Gunther and Wolff (1980), Sauer (1979).

The renewal approach results in perfectly independent cycle responses (unless we choose an approximate renewal state). We prefer independent responses because it simplifies the statistical analysis. Nevertheless the statistical analysis in the renewal approach involves some statistical problems because _ratio_ estimation is needed, as we show now. Consider the estimation of the expected waiting time in the steady-state of the M/M/1 queuing system. We shall use the following symbols:

μ : mean waiting time in steady-state ($\mu \equiv E(w_t)$ for $t \to \infty$).

n : (fixed) number of (independent) cycles.

L_i: (random) length of cycle i, i.e., number of customers in cycle i (i = 1,...,n).

W_{ij}: waiting time of customer j in cycle i (j = 1,...,L_i).

Consequently the total waiting time accumulated within cycle i is

$$y_i = \sum_{j=1}^{L_i} w_{ij} \qquad (6)$$

The traditional estimator of μ can be rewritten as follows:

$$\bar{w} = \sum_{t=1}^{N} w_t/N = \sum_{i=1}^{n} y_i / \sum_{i=1}^{n} L_i = \bar{y}/\bar{L} = \sum_i (\frac{L_i}{N}) (\frac{y_i}{L_i}) = \sum_i (\frac{L_i}{N}) \bar{w}_i \qquad (7)$$

which shows that the traditional estimator is identical to the cycle averages $\bar{w}_i$ weighted with the relative number of customers per cycle L_i/N. It can be proved that the true mean μ equals the following ratio of means: $\mu = E(y)/E(L)$. To estimate such a ratio a variety of estimators and confidence intervals is available. One obvious point estimator was shown in (7). However, we know that $E(\bar{y}/\bar{L}) \neq E(\bar{y})/E(\bar{L}) = E(y)/E(L)$, i.e., this "naive" estimator is biased. Fortunately, in large samples the naive estimator becomes unbiased. The confidence interval for this estimator is somewhat complicated because — although the cycles are independent — (7) showed that we do not have a simple average of independent observations. It can be proved (via the Central Limit Theorem applied to $y - \mu L$) that the following $(1-\alpha)$ confidence interval holds asymptotically:

$$\bar{w} \pm z^{\alpha/2} \cdot (\hat{\sigma} / \sqrt{n}) / \bar{L} \tag{8}$$

where

$$\hat{\sigma}^2 \equiv \hat{var}(y) - 2\bar{w} \cdot \hat{cov}(y,L) + (\bar{w})^2 \cdot \hat{var}(L) \tag{9}$$

so that the confidence interval becomes "better" (tighter) if

- we accept a higher α error (smaller $z^{\alpha/2}$),

- $\hat{\sigma}$ becomes smaller, i.e., the (estimated) variance of y or L becomes smaller, or their covariance becomes higher (an overshoot of y is compensated by an overshoot of L),

- n becomes higher (more independent cycles),

- $\bar{L}$ becomes higher (more customers per cycle).

The literature gives alternative point estimators and confidence intervals, but none seems to dominate the above procedure. One alternative estimator uses "jackknifing" and will be presented in the next section. See Crane and Lemoine (1977), Fishman (1978b).

Note that the sample size in the renewal approach is defined as the number of cycles, not as the number of individual customers. The renewal approach also solves the transient state problem (a serious problem in all other approaches): any renewal state may be selected for the initialization of the simulation run; no observations need to be thrown away. Selecting a different renewal state may yield longer cycles; for a given total computer time fewer cycles result but each cycle contains more information. Hence the efficiency of the renewal estimator does not depend on the renewal state selected. However, because of the asymptotic character of the confidence interval in (8), we prefer a renewal state resulting in many (possibly shorter) cycles.

The renewal approach applies to the estimation of the mean and other quantities such as proportions (or percentages), variances and quantiles; see the literature and Appendices 2 and 3.

A number of researchers has further developed the renewal approach during the past decade. They have investigated asymptotic results through analysis, and small-sample performance through simulation. Several types of simulation models have been analyzed in this way; see Section 8 on applications. The challenge for practitioners is now to discover the renewal property of their more complicated systems. In the mean time research continues, e.g., recently Heidelberger and Lewis (1981a) studied the detection and removal of small-sample bias and nonnormality of renewal estimators. Nonnormality may result in positively correlated estimators of the mean and variance, and this correlation may result in incorrect confidence intervals. We shall next present two very general approaches for diminishing nonnormality and bias respectively, where Heidelberger and Lewis (1981a) used more sophisticated regression and graphical techniques.

6. Nonnormality and bias: general techniques

Nonnormality may be reduced by combining the renewal approach with a simple kind of batching (this batching is also used in quantile estimation). Suppose we have available n cycles, resulting in the independent pairs y_i and L_i (i = 1,...,n). Let us divide these n cycles into m groups of cycles each group comprising a = n/m cycles. Per group we compute a point estimator using (7):

$$\bar{w}_g = \bar{y}_g / \bar{L}_g \qquad (g = 1, \ldots, m) \qquad\qquad (10)$$

where the numerator $\bar{y}_g$ is the average of group g; that group comprises "a" cycles, each cycle yielding a value y analogous to (6):

$$\bar{y}_g = \sum_{h=1}^{a} y_{gh}/a \text{ and } y_{gh} = \sum_{j=1}^{L_{gh}} w_{ghj} \qquad\qquad (11)$$

The denominator in (10) is the analogue of (11). Since the cycles give independently and identically distributed (i.i.d.) variables the group averages $\bar{w}_g$ of (10) are also i.i.d. Hence we can compute a confidence interval from the m group estimates $\bar{w}_g$ using the familiar t statistic with m - 1 degrees of freedom. Because the group averages $\bar{w}_g$ are i.i.d. the Central Limit Theorem explains why the overall average $\bar{\bar{w}} = \sum \bar{w}_g/m$ is more normally distributed than was the original point estimator $\bar{w}$ in (7). Besides, to the i.i.d. group estimators $\bar{w}_g$ we can apply non-parametric procedures such as the sign or rank test. One disadvantage of grouping or "batching" of cycle responses (or any other estimators outside the renewal analysis) is that the variance is estimated from fewer independent observations (m < n). Another disadvantage is that the small-sample bias of the point estimators $\bar{w}_g$ (and hence the bias of their average $\bar{\bar{w}}$) exceeds the bias of the original estimator $\bar{w}$. Also see Fishman (1978b, pp. 120-122), Heidelberger and Lewis (1981a), Kleijnen (1975, p. 501).

The small-sample bias of the group estimators $\bar{w}_g$ and of the original estimator $\bar{w}$ can be reduced through the jackknife technique. Jackknifing was proposed by Quenouille in 1949 as a general "trick" for

reducing bias; in 1969 Tukey emphasized that this technique results in a confidence interval even if the individual observations are dependent. We shall present the jackknife because it is quite often used in renewal analysis to reduce bias in short runs; moreover in other analyses the jackknife may also be a useful idea. So suppose we have an estimator $\hat{\theta}$ based on n independent observations x_j with $j = 1,\ldots,n$ (x may be multivariate). For instance, the estimator in renewal analysis is $\hat{\theta} \equiv \bar{y}/\bar{L}$. We divide the original sample into N groups of equal size $M = n/N$ where N is an integer exceeding the value one but possibly as big as n; see (12). We form a subsample by <u>deleting</u> one group from the N groups, and from the remaining $(N-1)M$ observations on x we compute the same estimator, denoted by $\hat{\theta}_i$ $(i = 1,\ldots,N)$. For instance, in renewal analysis we may delete a single cycle i and compute

$$\hat{\theta}_i = \sum_{j \neq i} y_j / \sum_{j \neq i} L_j \qquad (i = 1,\ldots,N = n) \tag{12}$$

Then the "pseudo values" J of the jackknife estimator are defined as

$$J_i = N \cdot \hat{\theta} - (N-1) \cdot \hat{\theta}_i \qquad (i = 1,\ldots,N) \tag{13}$$

Obviously, if $\hat{\theta}$ (and hence $\hat{\theta}_i$) were unbiased, then J would be unbiased. The <u>jackknife</u> estimator of θ is

$$\bar{J} = \sum_1^N J_i / N = N \cdot \hat{\theta} - (N-1) \cdot \bar{\hat{\theta}}_i \tag{14}$$

where $\bar{\hat{\theta}}_i \equiv \sum_1^N \hat{\theta}_i / N$. It can be shown that in many cases the jackknife estimator reduces possible bias in $\hat{\theta}$ from order n^{-1} to order n^{-2}. The

pseudovalues J might be treated as N i.i.d. variables, and an approximate confidence interval can then be based on the t statistic:

$$t \underset{\sim}{} \frac{\bar{J} - \theta}{s_J / \sqrt{N}} \quad \text{with} \quad s_J^2 = \frac{\sum\limits_1^N (J_i - \bar{J})^2}{N-1} \tag{15}$$

To make (15) hold better, we may transform the variable x before jack-knifing, e.g., we may jackknife log s rather than s itself. More details on the jackknife can be found in Cressie (1981) and Kleijnen (1975).

7. Other approaches

For the analysis of nonterminating simulations we discussed in detail: replicated runs, "independent" batches of fixed length, and the renewal approach. The literature shows some more approaches.

We may estimate the autocorrelation coefficients ρ_d among the individual observations x_t; see (2). For large lags (high d values) only a few observations are available and the estimators of ρ_d show large standard errors. Instead of estimating ρ_d we may estimate a transformation of ρ_d, i.e., we may resort to spectral analysis (involving the Fourier transformation of ρ_d). Unfortunately this analysis is mathematically sophisticated so that most practitioners hesitate to apply spectral analysis (alternatively, the practitioner may use a spectral analysis package as a black box). See Heidelberger and Welch (1981) and Fishman (1978b).

A different approach expresses the observation x_t as a moving average. This approach seems too sophisticated for most practitioners, and it involves several technical problems. See Fishman (1978b).

Note that originally (outside simulation) spectral analysis and the moving average representation were developed, not to obtain an estimator for the mean with its standard error, but to characterize the autocorrelation structure of the observations, i.e., this analysis tries to detect periodicities and considers the time series x_t as a composition of variables each with their own frequency of oscillation; see Box and Jenkins (1978), Fishman (1978b).

Recently a different approach has been proposed by Schruben (1982), based on the Wiener or Brownian Motion process; also see Appendix 1. Although this process involves a sophisticated theory, Schruben's analysis results in quite simple procedures. It seems too early to make definitive recommendations concerning this new technique.

In all approaches of this section there remains the problem of how to handle initialization effects. Renewal analysis is the only technique that eliminates the initialization problem. The estimation of variances and quantiles is discussed in Appendices 2 and 3.

8. Applications

In the preceding sections we presented a plethora of statistical techniques for the analysis of different situations. These situations may differ in many respects:

- terminating versus nonterminating (steady-state) simulations,

- different measures: mean, variance, quantiles, proportions,

- single versus multiple responses.

Most publications on applications of statistical techniques in simula-

tion, concern steady-state situations with a single response, namely the mean.

Many publications referenced in the preceding sections contain empirical results, obtained when applying one or more statistical techniques to simulation models. (See Crane and Lemoine (1977), Heidelberger and Lewis (1981a, 1981b), Heidelberger and Welch (1981), Sauer (1979).) These models usually represent queuing systems, ranging from the simple M/M/1 system to queuing networks (inspired by the modeling of computer systems). Such empirical results are needed because the statistical theory underlying the various techniques, usually gives asymptotical results.

Extensive experiments with the renewal analysis have further been conducted by Lavenberg and Sauer (1977). A confidence interval's relative width of 5% resulted in valid confidence intervals in almost all their experiments; but also "in many experiments larger relative widths were adequate"; a relative width larger than 5% results in smaller sample sizes so that asymptotic results may not apply. Iglehart and Shedler (1982) have extended and applied the renewal approach to models of closed queuing networks with priorities among job classes; these models are used in computer and communication systems analysis.

Experiments with batching were performed by Fishman (1978a) - but he applied the Von Neumann test to eight (or more) subruns whereas we recommended to take at least a hundred subruns because of the small power of the Von Neumann test. He found that the batching approach worked in M/M/1 queuing simulations with a traffic intensity λ as small

as 0.50; the higher λ the smaller the coverage. This lower coverage is explained by the underestimation of the variance of the batch averages caused by neglecting remaining autocorrelations among these averages.

Law and his associates Carson and Kelton performed a series of experiments with the three main techniques discussed in this paper (in terminating and non-terminating simulations, applying both fixed-sample and sequential sampling plans, with relative and absolute width of confidence intervals). All experiments were restricted to estimation of the mean. Their most recent results come from simulating the following two processes: the M/M/1 model with traffic load $\lambda = 0.8$, and a time-sharing computer system model with known analytical solution. From this series of experiments no statistical technique emerged as valid in all situations. For a summary of their results (and references) we refer to Law (1982) or Law and Kelton (1982, pp. 279-315). Other applications are surveyed in Kleijnen (1975, 1984).

If we determine confidence intervals for more than a single quantity, then we can use the Bonferroni inequality (i.e., the experimentwise error rate does not exceed the sum of the individual error rates α; see Miller 1966, 1981, and also Kleijnen, 1975a). Many authors are not aware of the issue at stake when making multiple inferences, and they simply use classical tests like the t test with classical α values like $\alpha = 0.05$. A publication explicity using the Bonferroni inequality, is Lavenberg and Slutz (1975) who simulated an automated computer-tape library.

Applications of <u>nonparametric</u> tests to simulation data are extremely rare. Remember that many studies showed that in simulation low coverage results as a consequence of the dependence between sample average and variance in non-normal distributions (i.e., distributions with a long tail to the right). Whether nonparametric procedures remove this low coverage, deserves more research; also see Heidelberger and Lewis (1981b).

In summary, if the number of "observations" (cycles, batches, replications) is large enough then valid confidence intervals result; in small samples the interval may miss the true mean with a probability exceeding α. Unfortunately analytical results for the various statistical procedures assume asymptotic normality; empirical results are limited to a few relatively simple simulation models. Recently several publications have discussed criteria for empirical research in this field; see Law (1982), Schriber and Andrews (1981).

The "applications" above concerned experiments by researchers on simulation methodology. There is another category of simulation users, namely researchers on non-simulation problems (e.g. queuing problems) who use simulation. Because of their scientific attitude we expect that these researchers are willing to apply statistical techniques in the analysis of their simulation data. For instance, Pinedo and Wolff (1982) applied renewal analysis in their simulation experiments with tandem queues, simulations being used to verify analytical approximations. In simulation experiments by real practitioners the statistical analysis is completely missing or is rudimentary, e.g., they use the batching ap-

proach with intuitively chosen subrun length. Our experience is that simulation practitioners can learn the statistical techniques of this survey without too much trouble.

9. Summary:

Sometimes simulations are nonterminating and the analyst is interested in the steady-state mean. We cannot recommend replication of runs because each run shows initialization effects. In practice it is customary to partition the single, long run into subruns of fixed length. We recommend to test whether the subrun responses are indeed independent, using at least a hundred subruns (power of Von Neumann test). Renewal analysis solves the initialization problem completely, and gives perfectly independent responses. For more complicated non-terminating simulations we may use an approximate renewal state. Other approaches such as spectral analysis seem too sophisticated for practical use.

Appendix 1. The Schruben-Singh-Tierney test for initialization bias

Schruben et al. (1980) proposed the following procedure for testing whether there is bias in a time series (if there is "obvious" bias in the beginning part of the output series, the procedure can be applied to the remaining truncated output).

(i) Consider the sequence of differences (d_k) between the average ($\bar{w}$) of the first k observations, $\bar{w}_k = \sum_1^k w_i/k$, and the overall average $\bar{w}_N = \sum_1^N w_i/N$:

$$d_k = \bar{w}_k - \bar{w}_N \qquad (k = 0,1,\ldots,N) \qquad (1.1)$$

with $d_0 \equiv 0$ and obviously $d_N = 0$. (Note that these differences lead to the so-called Brownian bridge process; d is also related to the CUSUM-tests in quality control; the running averages $\bar{w}_k$ are compared to the final best estimate $\bar{w}_N$.)

(ii) Estimate σ^2, the asymptotic variance of $(\bar{w}_N-\mu) \surd N$, in other words, var $(\bar{w}_N) = \sigma^2/N$ for large N. Several approaches are available for estimating var $(\bar{w}_N)$, e.g., divide the N observations w_i into n batches or replicate the whole time series w_i n times; see Sections 3 and 4. The estimation of σ^2 results in $\hat{\sigma}^2$ with d degrees of freedom. It is wise to estimate σ^2 from the last half of the runs because of initialization effects. Hence in the batching approach the degrees of freedom become $(n/2) - 1$.

(iii) A likelihood ratio test leads to the t statistic:

TABLE 1

Replication of Time Series

Replication	Time series					Variance estimator
1	x_{11}	x_{12} $\cdots$	x_{1t} $\cdots$	x_{1T}		$\hat{\sigma}^2_1$
2	x_{21}	x_{22} $\cdots$	x_{2t} $\cdots$	x_{2T}		$\hat{\sigma}^2_2$
.	.	.	.	.		.
.	.	.	.	.		.
.	.	.	.	.		.
i	x_{i1}	x_{i2}	x_{it} $\cdots$	x_{iT}		$\hat{\sigma}^2_i$
.	.	.	.	.		.
.	.	.	.	.		.
.	.	.	.	.		.
n	x_{n1}	x_{n2} $\cdots$	x_{nt} $\cdots$	x_{nT}		$\hat{\sigma}^2_n$
Variance estimator	s^2_1	s^2_2 $\cdots$	s^2_t $\cdots$	s^2_T		

$$t_d = (\frac{45}{N \cdot \hat{\sigma}^2})^{\frac{1}{2}} \sum_{k=1}^{N} d_k \cdot \frac{k}{N} \cdot (1 - \frac{k}{N}) \qquad (1.2)$$

(iv) Reject the hypothesis of no initialization bias if t is signifi-
cant. If the sign of possible initialization bias is known then a one-
sided test is appropriate; otherwise a two-sided test is in place. For
instance, in an M/M/1 queuing simulation started in the empty state, we
expect that $\mu_1 < \mu_2 < \ldots$ so that d_k tends to be negative and hence t
becomes significantly negative.

Appendix 2. Variance estimation

Suppose we wish to the estimate the variance σ_x^2 in the following
equation where the last two equalities hold for large t values ($t \rightarrow \infty$):

$$var(x_t) \equiv \int_{-\infty}^{\infty} (x - \mu_t) \cdot f_t(x) \, dx = \int_{-\infty}^{\infty} (x - \mu) \cdot f(x) \, dx = \sigma_x^2 \qquad (2.1)$$

If we replicate the simulation run n times we obtain an n by T array
with x_{it} (i = 1,...,n and t = 1,...,T); see Table 1. First consider
only the last observation of each run: x_{iT}. Then an unbiased estimator
of σ_x^2 is:

$$s_T^2 = \sum_{i=1}^{n} (x_{iT} - \bar{x}_T)^2 / (n-1) \qquad (2.2)$$

This is the simplest estimator but its degrees of freedom are only n-1.
We can also generalize (2.2) replacing T by t, resulting in the depen-
dent estimators s_t^2 with t = T, T-1, T-2,... (This dependence can be

illustrated as follows. Suppose the first replication yields a relatively high $x_{1(T-1)}$; then the autocorrelation among the x_t implies that x_{1T} tends to be high too. The extreme values of $x_{1(T-1)}$ and x_{1T} inflate both variance estimators s_{T-1}^2 and s_T^2 so that these estimators show positive correlation.) As t becomes smaller (t → 1) initialization bias tends to be more serious, and therefore we delete the warming-up period. The resulting estimators $s_{t'}^2$ (t' = T, T-1,...;t' ≫ 1) remain dependent. To obtain the standard error of the (unbiased) average of these $s_{t'}^2$, we can apply the <u>batching</u> approach to the time series $s_{t'}^2$ (or one of the other approaches briefly mentioned in Section 7).

Serious underestimation of σ_x^2 would result if we estimated σ_x^2 from batches of observations on x_t: the autocorrelation among x_t means that little variation exists within short subruns. Hence each batch j (j = 1,...,m) tends to underestimate σ_x^2. However, taking the deviation of x_t not from its corresponding batch average $\bar{x}_j$ but from the overall average $\bar{\bar{x}} = \sum_1^m \bar{x}_j/m = \sum_1^T x_t/T$ results in the last column of Table 1:

$$\hat{\sigma}_i^2 = \sum_{t=1}^T (x_{it} - \bar{\bar{x}}_i)^2/(T-1) \qquad (i = 1,...,n) \qquad (2.3)$$

Note that in the estimator $\hat{\sigma}_i^2$ dependent observations x_{it} occur, whereas the estimator s_t^2 in the last row of Table 1 is a classical estimator. More research on variance estimation is necessary; also see Welch (1983). But remember that renewal analysis permits straightforward estimation of σ_x^2 from a single run.

Note that a system is called <u>ergodic</u> if the estimators computed from a single run are equivalent to the estimators computed from replicated runs. Consider the following (artificial) example of a non-ergodic system, taken from Mihram (1972, p. 448):

$$x_t = 1 \text{ for all t, if the toss of a die is 1 or 2}$$
$$" = 0 \text{ " } \qquad " \qquad " \qquad 3 \text{ or } 4$$
$$" = -1 \text{ " } \qquad " \qquad " \qquad 5 \text{ or } 6 \qquad (2.4)$$

Consequently $\hat{\sigma}_i^2 = 0$ and $E(\hat{\sigma}_i^2) = 0$ for all i, whereas $E(s_t^2) = 2/3$ for all t. A more realistic example involving an IBM business game is presented in Kleijnen (1980, pp. 157-186). Many stationary processes, however, will be ergodic, e.g., processes with the renewal property also have the ergodicity property.

Appendix 3. Quantile estimation

The pth quantile x_p in a steady-state system is defined by the following equation, where the last equality holds for large t values:

$$p = P(X_t < x_p) = \int_{-\infty}^{x_p} f_t(x) \, dx = \int_{-\infty}^{x_p} f(x) \, dx \qquad (3.1)$$

For regenerative simulations three different techniques are available: Iglehart (1976), Seila (1976), and Moore (1980). We present only Seila's technique, since the other two techniques are quite complicated (nevertheless, Iglehart's technique was applied in a computer center case study - see Keyzer et al. 1981 - and Moore's technique may show better

statistical performance). Seila groups the n cycles into m groups of equal size a; also see (10). Each group yields the usual estimator of x_p. Arrange the observations x_i in increasing order, i.e., obtain the order statistics $x_{(i)}$. The sample quantile is then $x_{(p.n+1)}$. Hence, if L_i denotes the length of cycle i (i = 1,...,n) and B denotes the number of customers per group $(B_1 = \sum_{i=1}^{a} L_i, B_2 = \sum_{a+1}^{2a} L_i,$ etc.) then we get $\hat{x}_p = x_{(B.p + 1)}$. Because of the renewal property, cycles are independent and hence groups of cycles are independent. Hence the m groups yield m independent estimators of x_p. The average of these m observations results in a point estimator of x_p; the confidence interval may be based on the t statistic with m-1 degrees of freedom. Since the usual estimator of x_p is only asymptotically unbiased, Seila applies jackknifing to reduce small-sample bias. (He divides the group into two subgroups, each comprising a/2 cycles: N = 2 in (12) through (15); we conjecture that taking N = a results in better statistical results at the expense of more computer time.) Note that instead of using the parametric t statistic we may compute a nonparametric confidence interval; see Conover (1971, pp. 110-111, 215-222).

Seila (1976) applied his procedure to the M/M/1 system with traffic intensity λ between 0.5 and 0.9. He found coverages slightly larger than the nominal 1-α value. The price paid is slightly longer confidence intervals: jackknifing inflated the variance. The estimation of quantiles requires much larger sample sizes than does the estimation of means; typically where mean estimation requires a hundred cycles quantile estimation takes a thousand cycles. To reduce bias large batch sizes (say, at least a hundred cycles per batch) are recommended, even

when jackknifing. We could add the following heuristic to Seila's procedure: compute the point estimate from the whole run (n cycles) but compute the confidence interval from the groups (without jackknifing?); if the point estimator lies "near" one end of the confidence interval then simulate more cycles. The performance of such a heuristic remains to be investigated.

Seila (1982) proposed to use his procedure not only with the regenerative approach but also with the replication or batching approach. In the latter two approaches the initial part of the run may be deleted. We add that a test for the independence of the batch quantiles has little power unless we have a hundred batches (with a regenerative system such a test is not needed since the cycle responses are known to be independent). Seila's grouping approach (besides spectral analysis) was applied by Heidelberger and Lewis (1981b, pp. 35-36) with acceptable results.

In quantile estimation we have a storing and sorting problem. With positively autocorrelated observations larger sample sizes are needed, and this computer problem becomes even more serious. For computer algorithms we refer to Sedgewick (1978); for statistical issues see Heidelberger and Lewis (1981b); also see Kleijnen (1984).

References

Adlakha, V.G. and G.S. Fishman (1982). Starting and stopping rules for simulations using a priori information. European Journal Operational Research, 10, no. 4: 379-394.

Box, G.E.P., and G.M. Jenkins (1978). Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco, second edition.

Conover, W.J. (1971). Practical Non-parametric Statistics. John Wiley & Sons, Inc., New York.

Crane, M.A. and A.J. Lemoine (1977). An Introduction to the Regenerative Method for Simulation Analysis. Springer-Verlag, Berlin.

Cressie, N. (1981). Transformations and the jackknife. Journal Royal Statistical Society, Series B, 43, no. 2: 177-182.

Fishman, G.S. (1978a). Grouping observations in digital simulation. Management Science, 24, no. 5: 510-521.

Fishman, G.S. (1978b). Principles of Discrete Event Simulation. John Wiley & Sons, Inc., New York.

Gunther, F.L. and R.W. Wolff (1980). The almost regenerative method for stochastic system simulations. Operations Research, 28, no. 2: 375-386.

Heidelberger, P. and P.A.W. Lewis (1981a). Regression-adjusted estimates for regenerative simulations, with graphics. Communications ACM. 24, no. 4: 260-27 .

Heidelberger, P. and P.A.W. Lewis (1981b). Quantile Estimation in Dependent Sequences. IBM Research Center, Yorktown Heights (New York).

Heidelberger, P. and P.D. Welch (1981). Adaptive spectral methods for simulation output analysis. IBM Journal of Research and Development, 25, no. 6: 860-876.

Iglehart, D.L. (1976). Simulating stable stochastic systems, VI: quantile estimation. Journal Association Computing Machinery, 23, no. 2, pp. 347-360.

Iglehart, D.L. and G.S. Shedler (1982). Comparison of Regenerative Simulation Methods for Networks of Queues with General Service Times. RJ 3415 (40681), IBM Research Laboratory, San Jose (California).

Janssens, G.K. (1982). References on stochastic processes, European Journal of Operational Research. 10, no. 4: 421-422.

Leyzer, F., J. Kleijnen, E. Mullenders and A. Van Reeken (1981). Optimization of priority class queues, with a computer center case study. American Journal Mathematical and Management Sciences, 1, no. 4: 341-358.

Kleijnen, J.P.C. (1975). Statistical Techniques in Simulation. Volumes I and II. Marcel Dekker, Inc., New York. (Russian translation: Publishing House "Statistics", Moscow, 1978.)

Kleijnen, J.P.C. (1980). Computers and Profits; Quantifying Financial Benefits of Information. Addison-Wesley Publishing Company, Reading (Massachusetts).

Kleijnen, J.P.C. (1984). Statistical tools for simulation practitioners. Marcel Dekker, Inc., New York.

Kleijnen, J.P.C., R. Van der Ven and B. Sanders (1982). Testing independence of simulation subruns: a note on the power of the Von Neumann test. European Journal of Operational Research, 9, no. 1: 92-93.

Lavenberg, S.S. and C.H. Sauer (1977). Sequential stopping rules for the regenerative method of simulation. IBM Journal of Research and Development, 21, no. 6: 545-558.

Lavenberg, S.S. and D.R. Slutz (1975). Regenerative simulation of a queuing model of an automated tape library. IBM Journal Research and Development, 19: 463-475.

Law, A.M. (1982). Statistical Analysis of Simulation Output Data: the State of the Art. College of Business, The University of Arizona, Tucson.

Law, A.M. and W.D. Kelton (1982). Simulation Modeling and Analysis. McGraw-Hill Book Company, New York.

Mihram, G.A. (1972). Simulation: Statistical Foundations and Methodology. Academic Press, New York.

Miller, R.G. (1966). Simultaneous Statistical Inference. McGraw-Hill Book Company, New York. Second edition. Revised edition: Springer-Verlag, New York, 1981.

Moore, L.W. (1980). Quantile Estimation in Regenerative Processes. Ph.D. dissertation, Department of Statistics, University of North Carolina, Chapel Hill.

Pinedo, M. and R.W. Wolff (1982). A comparison between tandem queues with dependent and independent service times. Operations Research, 30, no. 3: 464-479.

Sauer, C.H. (1979). Confidence intervals for queueing simulations of computer systems. Performance Evaluation Review, 8, no. 1 and 2: 45-55.

Schmeiser, B.W. (1982). Batch size effects in the analysis of simulation output. Operations Research, 30, no. 3: 556-568.

Schriber, T.J. and R.W. Andrews (1981). A conceptual framework for research in the analysis of simulation output. Communications ACM, 24, no. 4: 218-232.

Schruben, L.W. (1982). Confidence Interval Estimation from Standardized Simulation Output. School of Operation Reseach and Industrial Engineering, Cornell University, Ithaca (New York).

Schruben, L.W., H. Singh and L. Tierney (1980). A Test of Initialization Bias Hypotheses in Simulation Output. School of Operation Research and Industrial Engineering, Cornell University, Ithaca (New York).

Sedgewick, R. (1978). Implementing QUICKSORT programs. Communications ACM. 21, no. 10: 847-857.

Seila, A.F. (1976). Quantile Estimation Methods for Discrete Event Simulations of Regenerative Systems. Ph.D. dissertation, Operations Research and Systems Analysis, University of North Carolina, Chapel Hill.

Seila, A.F. (1982). Percentile Estimation in Discrete Event Simulation. College of Business Administration, University of Georgia, Athens. (Part of this paper already published in Management Science, 28, no. 5: 573-581.)

Welch, P.D. (1983). The statistical analysis of simulation results. Computer Performance Modeling Handbook, edited by S.S. Lavenberg, Academic Press, New York.

Wilson, J.R. and A.A.B. Pritsker (1978a). A survey of research on the simulation startup problem. Simulation, 31, no. 2: 55-58.

Wilson, J.R. and A.A.B. Pritsker (1978b). Evaluation of startup policies in simulation experiments. Simulation, 31, no. 3: 79-89.

Woodside, C.M. and B. Pagurek (1979). An algorithm for computing serial correlations of times in GI/G/1 queues with rational arrival processes. Management Science, 25, no. 1: 54-63.

IN 1982 REEDS VERSCHENEN:

107 Aart J. de Zeeuw
Hierarchical decentralized optimal control in econometric
policy models.

108 Arie Kapteyn en Tom Wansbeek
Identification in Factor Analysis.

109 G. van der Laan en A.J.J. Talman
Simplical Algorithms for finding Stationary Points, a unifying
description

110 Pieter Boot
Ekonomiese betrekkingen tussen Oost en West Europa.

111 B.B. van der Genugten
The asymptotic behaviour of the estimated generalized least squares
method in the linear regression model.

112 Jack P.C. Kleijnen / Anton J. van Reeken
Principles of computer charging in a university-like organization.

113 H. Tigelaar.
The informative sample size for dynamic multiple equation systems with
moving average errors.

114 Drs. H.G. van Gemert, Dr. R.J. de Groof en Ir. A.J. Markink
Sektorstruktuur en Economische Ontwikkeling

115 P.H.M. Ruys
The tripolar model: a unifying approach to change.

116 Aart J. de Zeeuw
Policy solutions for Mini-Interplay, a Linked Model for two Common
Market Countries.

IN 1982 REEDS VERSCHENEN (vervolg):

117 F.J.M. van Doorne en P.H.M. Ruys
    Die Struktur einer Sprechhandlung in Habermas' Forschungsprogramm.
    Formale Analyse mit den Mitteln des tripolaren Modells.

118 A.J.J. Talman en G. van der Laan
    Simplicial Approximation of Solutions to the Non-Linear Complementarity
    Problem

119 Pieter Boot
    The East German Planning System Reconsidered

120 J.P.C. Kleijnen
    Regression Metamodel Summarization of Model Behaviour

121 P.G.H. Mulder en A.L. Hempenius
    Evaluating life time in a competing risks model for a chronic desease.

122 A.J.J. Talman en G. van der Laan
    From Fixed Point to Equilibrium

123 J.P.C. Kleijnen
    Design of simulation experiments.

124 H.L. Theuns en A.M.L. Passier-Grootjans
    Internationaal toerisme; een gids in de algemene basisliteratuur en
    het bronnenmateriaal.

125 J.H.F. Schilderinck
    Part I: Imports and Exports, Sub-divided by Countries Aggregated
    According the Branches of the European Community Interregional Input-
    Outputtables 1959, 1965, 1970 and 1975.

IN 1983 REEDS VERSCHENEN:

126   Harry H. Tigelaar
      Identification of noisy linear systems with multiple arma inputs.